WEIGHT AND RANK OF MATRICES OVER FINITE FIELDS

THERESA MIGLER [*], KENT E. MORRISON[†], AND MITCHELL OGLE[†]

Abstract. Define the weight of a matrix to be the number of non-zero entries. One would like to count m by n matrices over a finite field by their weight and rank. This is equivalent to determining the probability distribution of the weight while conditioning on the rank. The complete answer to this question is far from finished. As a step in that direction this paper finds a closed form for the average weight of an m by n matrix of rank k over the finite field with q elements. The formula is a simple algebraic expression in m, n, k, and q. For rank one matrices a complete description of the weight distribution is given and a central limit theorem is proved.

Key words. Random Matrices, Finite Fields, Weight, Rank.

AMS subject classifications. 15A52 (primary), 15A03, 15A33, 60C05 (secondary).

1. Introduction. For an $m \times n$ matrix A over the finite field $\mathbb{F}_q$ the **weight** of A, denoted $\operatorname{wt} A$, is the number of non-zero entries. In the Hamming metric on matrices it is the distance from A to 0.

There is some relationship between the rank and the weight of a matrix. For example, if $\operatorname{wt} A = 1$, then $\operatorname{rk} A = 1$, and if $\operatorname{rk} A = k$, then $\operatorname{wt} A \geq k$. On the other hand, there are matrices of rank one and maximal weight mn, such as a matrix with every entry a one. In this article we determine the average weight of rank k matrices in terms of k, m, n, and q. Without fixing the rank, the average weight of $m \times n$ matrices is $mn(1 - \frac{1}{q})$ and the weight has a binomial distribution. However, the full probability distribution of the weight for matrices of rank k is yet to be determined.

The tools are those of elementary combinatorics and linear algebra. Nothing special is used from the theory of finite fields other than the understanding that the fundamental ideas of linear algebra work over all fields and not just the real or complex numbers.

We need a modest amount of background material. We use three basic formulas.

Formula 1.1. *The number of ordered k-tuples of linearly independent vectors in ${\mathbb{F}_q}^n$ is*

$$(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{k-1})$$

Proof. The first vector is any non-zero vector and each succeeding vector must avoid the span of the previous vectors. □

Formula 1.2. *The number of k-dimensional subspaces of ${\mathbb{F}_q}^n$ is given by the q-binomial coefficient*

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{k-1})}{(q^k - 1)(q^k - q)(q^k - q^2) \cdots (q^k - q^{k-1})}$$

[*]Department of Mathematics, California Polytechnic State University, San Luis Obispo, California 93407, USA (tmigler@calpoly.edu, kmorriso@calpoly.edu, mogle@calpoly.edu).

Proof. The numerator is the number of bases of all k-dimensional subspaces, while the denominator is the number of bases of any given subspace. □

FORMULA 1.3. *The number of $m \times n$ matrices of rank k is*

$$\begin{aligned} &\begin{bmatrix} m \\ k \end{bmatrix}_q (q^n - 1)(q^n - q) \cdots (q^n - q^{k-1}) \\ = &\begin{bmatrix} n \\ k \end{bmatrix}_q (q^m - 1)(q^m - q) \cdots (q^m - q^{k-1}) \\ = &\frac{(q^m - 1)(q^m - q) \cdots (q^m - q^{k-1})(q^n - 1)(q^n - q) \cdots (q^n - q^{k-1})}{(q^k - 1)(q^k - q)(q^k - q^2) \cdots (q^k - q^{k-1})} \end{aligned}$$

Proof. For a fixed k-dimensional subspace $W \subset \mathbb{F}_q{}^m$, the number of matrices with W as the column space is equal to the number of $k \times n$ matrices of rank k. Such a matrix is given by the k linearly independent row vectors of length n. The number of those is given by Formula 1. The number of k-dimensional subspaces of $\mathbb{F}_q{}^m$ is $\begin{bmatrix} m \\ k \end{bmatrix}_q$ and the product is the number of rank k matrices given in the first line. By the same reasoning, the second line counts the number of $n \times m$ matrices of rank k, which is the same. □

A special case of Formula 3 is worth noting. The number of invertible $n \times n$ matrices is

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1})$$

2. Average Weight. The average weight of a rank k matrix is the sum of the average weights of the entries, and the average weight of the ij entry is the probability that the entry is not zero:

$$\mathbf{E}(\mathrm{wt}\, A) = \sum_{i,j} \mathbf{P}(a_{ij} \neq 0)$$

THEOREM 2.1. *The probability that $a_{ij} \neq 0$ for a rank k matrix A is the same for all i and j.*

Proof. The probability that the ij entry is not zero is the quotient whose numerator is the number of matrices A of rank k with $a_{ij} \neq 0$, and whose denominator is the number of matrices of rank k. Consider the map on the $m \times n$ matrices that switches rows 1 and i and switches columns 1 and j. This map preserves rank and gives a bijection between the subset of matrices of rank k with a non-zero in the 1,1 location and the subset of matrices of rank k with a non-zero in the i, j location. Thus, $\mathbf{P}(a_{ij} \neq 0) = \mathbf{P}[a_{11} \neq 0)$. □

Call this common value **the average weight per entry**. Now we focus on the upper left corner of the matrices of rank k. Our analysis depends on the reduced row echelon form. We recall the definition [1].

DEFINITION 2.2. A rectangular matrix is in *row echelon form* if it has the following three properties:

1. All non-zero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zero.

If a matrix in echelon form satisfies the following additional conditions, then it is in *reduced row echelon form*:

4. The leading entry in each nonzero row is 1
5. Each leading 1 is the only non-zero entry in its column.

The $k \times n$ matrices in reduced row echelon form correspond bijectively with the k-dimensional subspaces of ${\mathbb{F}_q}^n$. The rows of the matrix give a basis of the corresponding subspace. When an $m \times n$ matrix A is reduced to reduced row echelon form by row operations, the result is an $m \times n$ matrix whose first k rows form a basis of the row space of A. Let R be the $k \times n$ matrix consisting of the k non-zero rows of the reduced form. Then A and all matrices with the same row space can be constructed from R by multiplying R on the left by an $m \times k$ matrix C of rank k. The matrix C is unique. This gives a factorization of A as $A = CR$. In terms of the associated linear maps, A is a linear map from ${\mathbb{F}_q}^n$ to ${\mathbb{F}_q}^m$, which factors into a surjective map onto ${\mathbb{F}_q}^k$ followed by an injective map from ${\mathbb{F}_q}^k$ to ${\mathbb{F}_q}^m$. Recall that knowing the row space of a matrix is equivalent to knowing the kernel of the associated linear map. Thus, when the reduced matrix R is held fixed and C is varied, the product CR gives all maps with the same row space (i.e. same kernel).

Theorem 2.3. *For $m \times n$ matrices of rank k, the average weight per entry is*

$$\frac{\left(1 - \frac{1}{q}\right)\left(1 - \frac{1}{q^k}\right)}{\left(1 - \frac{1}{q^m}\right)\left(1 - \frac{1}{q^n}\right)}$$

Proof. We consider the random selection of a rank k matrix such that each such matrix is equally probable. With the factorization $A = CR$, this can be done by selecting C uniformly from all $m \times k$ matrices of rank k and by selecting R independently from among all reduced row echelon matrices, which is the same as selecting the row space uniformly from all k-dimensional subspaces of ${\mathbb{F}_q}^n$. The upper left corner of A is $a_{11} = c_{11}r_{11} + c_{12}r_{21} + \cdots + c_{1k}r_{k1}$. But because R is in reduced row echelon form, the first column of R is either all zeros or has a leading 1 followed by zeros. Thus, $a_{11} = c_{11}r_{11}$. In order for a_{11} to be non-zero, both r_{11} and c_{11} must be non-zero. Since the selection of R is independent of the selection of C,

$$\mathbf{P}(a_{11} \neq 0) = \mathbf{P}(c_{11} \neq 0)\mathbf{P}(r_{11} \neq 0)$$

The columns of C are k linearly independent vectors of length m and so the first column is not the zero vector. That means there are $q^m - 1$ possible first column vectors. There are $(q-1)$ choices for $c_{11} \neq 0$ and q^{m-1} choices for the remaining entries of the first column. Therefore,

$$\mathbf{P}(c_{11} \neq 0) = \frac{(q-1)q^{m-1}}{q^m - 1} = \frac{q^m - q^{m-1}}{q^m - 1}$$

Now $r_{11} = 0$ or 1, and $r_{11} = 0$ when the row space of R contains nothing in the direction of the vector $(1, 0, 0, \ldots, 0)$, which is to say that the row space is contained in the $(n-1)$-dimensional space $\{(0, x_2, \ldots, x_n)\}$. Therefore,

$$\mathbf{P}(r_{11} = 0) = \frac{\begin{bmatrix} n-1 \\ k \end{bmatrix}_q}{\begin{bmatrix} n \\ k \end{bmatrix}_q}$$

$$\mathbf{P}(r_{11} \neq 0) = 1 - \frac{\begin{bmatrix} n-1 \\ k \end{bmatrix}_q}{\begin{bmatrix} n \\ k \end{bmatrix}_q}$$

Using Formula 2 one easily obtains

$$\frac{\begin{bmatrix} n-1 \\ k \end{bmatrix}_q}{\begin{bmatrix} n \\ k \end{bmatrix}_q} = \frac{q^{n-k} - 1}{q^n - 1}$$

Putting these results together we have

$$\begin{aligned} \mathbf{P}(a_{11} \neq 0) &= \frac{q^m - q^{m-1}}{q^m - 1} \left(1 - \frac{q^{n-k} - 1}{q^n - 1}\right) \\ &= \frac{(q^m - q^{m-1})(q^n - q^{n-k})}{(q^m - 1)(q^n - 1)} \\ &= \frac{\left(1 - \frac{1}{q}\right)\left(1 - \frac{1}{q^k}\right)}{\left(1 - \frac{1}{q^m}\right)\left(1 - \frac{1}{q^n}\right)} \qquad \square \end{aligned}$$

With this result we have a clear picture of the effect of the parameters k, m, and n on the average weight. The factor $1 - 1/q$ is the average weight per entry without the rank condition, in which case the matrix size does not matter. Note that with m and n fixed, it is more likely for an entry to be non-zero for matrices of larger rank, an intuitively plausible result because both weight and rank are some measure of distance from the zero matrix. Also, one can see that as m, n, and k simultaneously go to infinity, the probability approaches $1 - 1/q$, which is again the unconditioned probability. For invertible matrices of size n (i.e. $k = m = n$) the probability of a non-zero entry is

$$\frac{1 - \frac{1}{q}}{1 - \frac{1}{q^n}}$$

3. Weight of Rank One Matrices. For the matrices of rank one a more complete analysis of the weight distribution is possible. In this case C is a non-zero column vector of length m and R is a non-zero row vector of length n whose leading non-zero entry is 1. The rank one matrix $A = CR$ is given by $a_{ij} = c_i r_j$, and so the weight of A is the product of the weights of C and R. The weight of C has a binomial

distribution conditioned on the weight being positive (the entries of C cannot all be zero)

$$\mathbf{P}(\mathrm{wt}\, C = \mu) = \frac{\binom{m}{\mu}(q-1)^{\mu} q^{m-\mu}}{(q^m - 1)}$$

Likewise for R the weight distribution is given by

$$\mathbf{P}(\mathrm{wt}\, R = \nu) = \frac{\binom{n}{\nu}(q-1)^{\nu} q^{n-\nu}}{(q^n - 1)}$$

(To select a random R, choose a random non-zero vector of length n and then scale it to make the leading non-zero entry 1. The scaling does not change the weight.)

The weight on rank one matrices is the product of these two binomial random variables, each conditioned to be positive.

$$\mathbf{P}(\mathrm{wt}\, A = \omega) = \sum_{\mu\nu=\omega} \mathbf{P}(\mathrm{wt}\, C = \mu)\mathbf{P}(\mathrm{wt}\, R = \nu) \tag{3.1}$$

$$= \sum_{\mu\nu=\omega} \binom{m}{\mu}\binom{n}{\nu} \frac{(q-1)^{m+n-\mu-\nu} q^{\mu+\nu}}{(q^m-1)(q^n-1)} \tag{3.2}$$

Not all weights between 1 and mn occur for rank 1 matrices since the weight is a product with one factor no greater than m and the other factor no greater than n. Plots of actual probability distributions show spikes and gaps. Plots of cumulative distributions are smoother and lead us to expect a limiting normal distribution. See Figures 1 and 2.

Theorem 3.1. *As m or n goes to infinity, the weight distribution of rank one matrices approaches a normal distribution.*

Proof. The weight random variable for rank one matrices of size $m \times n$ is the product of independent binomial random variables conditioned on being positive. Define $W = XY$, where $X = \sum_{1\le i\le m} X_i$, $Y = \sum_{1\le j\le n} Y_j$, and X_i and Y_j are independent Bernoulli random variables with probability $1/q$ of being 0. Then W is the sum of m independent identically distributed random variables $X_i Y$. Conditioning W on $W > 0$ is the weight of rank one matrices. By the Central Limit Theorem the distribution of W converges, as $m \to \infty$, to a normal distribution after suitable scaling. Now conditioning on W being positive does not change this result because the probability that $W > 0$ is $1 - q^{-m}$, which goes to 1 as $m \to \infty$. ◻

Now to compute the mean and variance of the weight, let $W = XY$ as before but without conditioning on X or Y being positive. Then $\mathbf{E}(W) = mn(1 - 1/q)^2$ and

$$\mathbf{E}(W^2) = \mathbf{E}(X^2Y^2) = \mathbf{E}\left(\left(\sum_i X_i\right)^2 \left(\sum_j Y_j\right)^2\right)$$

Expanding and using the independence of the random variables X_i, Y_j and the fact

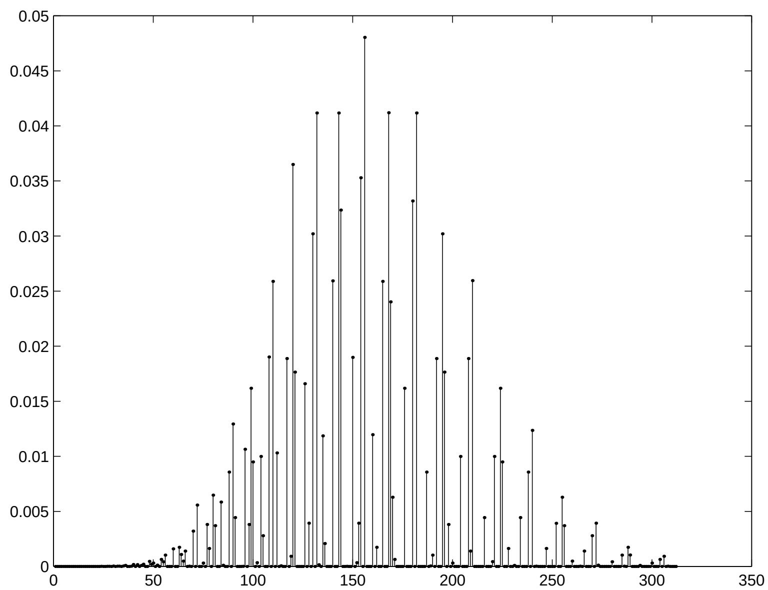

FIG. 3.1. *Distribution for the weight of rank 1 matrices, $m = n = 25$, $q = 2$.*

that $X_i^2 = X_i$ and $Y_j^2 = Y_j$, we get

$$\begin{aligned}\mathbf{E}(W^2) = mn\left(1-\frac{1}{q}\right)^2 + mn(m+n-2)\left(1-\frac{1}{q}\right)^3 \\ + m(m-1)n(n-1)\left(1-\frac{1}{q}\right)^4\end{aligned}$$

The variance of the weight is

$$\begin{aligned}\operatorname{var}(W|W>0) &= \mathbf{E}(W^2|W>0) - \mathbf{E}(W|W>0)^2 \\ &= \frac{\mathbf{E}(W^2)}{\mathbf{P}(W>0)} - \left(\frac{\mathbf{E}(W)}{\mathbf{P}(W>0)}\right)^2\end{aligned}$$

Furthermore

$$\mathbf{P}(W>0) = \mathbf{P}(X>0)\mathbf{P}(Y>0) = (1-\frac{1}{q^m})(1-\frac{1}{q^n})$$

Combining these expressions we get

$$\operatorname{var}(W|W>0) = \\ \frac{mn\left(1-\frac{1}{q}\right)^2 + mn(m+n-2)\left(1-\frac{1}{q}\right)^3 + m(m-1)n(n-1)\left(1-\frac{1}{q}\right)^4}{\left(1-\frac{1}{q^m}\right)\left(1-\frac{1}{q^n}\right)}$$

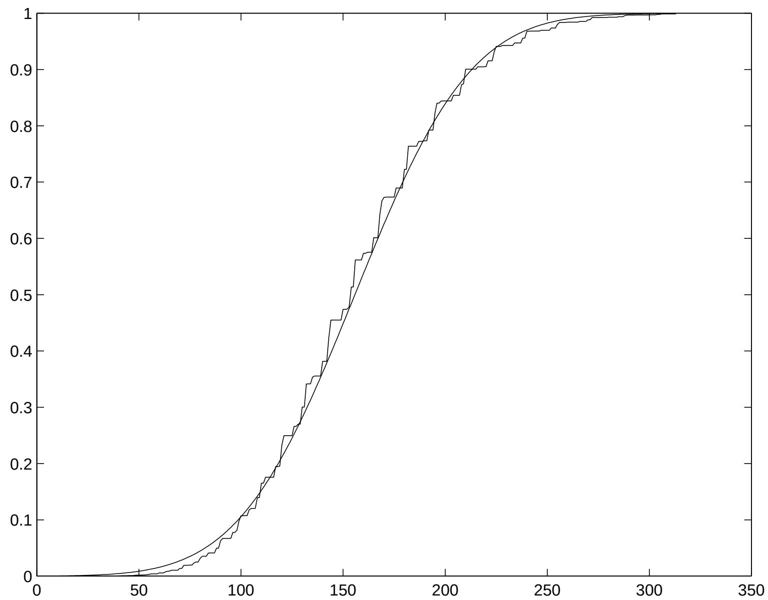

FIG. 3.2. *Cumulative frequency distribution for the weight of rank 1 matrices, $m = n = 25$, $q = 2$. The smooth curve is the normal cdf with the same mean (≈ 156.25) and standard deviation (≈ 44.63).*

$$- \frac{\left(mn\left(1-\frac{1}{q}\right)\right)^2}{\left(1-\frac{1}{q^m}\right)^2\left(1-\frac{1}{q^n}\right)^2}$$

We get a good approximation to the variance of the weight when m and n using the unconditioned W, essentially the factors in the denominator by 1. Thus,

$$\begin{aligned}
\operatorname{var}(W) &= \mathbf{E}(W^2) - \mathbf{E}(W)^2 \\
&= mn\left(1-\frac{1}{q}\right)^2 + mn(m+n-2)\left(1-\frac{1}{q}\right)^3 \\
&\quad + m(m-1)n(n-1)\left(1-\frac{1}{q}\right)^4 - \left(mn\left(1-\frac{1}{q}\right)\right)^2
\end{aligned}$$

which can be simplified to give

$$\begin{aligned}
\operatorname{var}(W) = mn(1-n-m)\left(1-\frac{1}{q}\right)^4 \\
+ mn(n+m-2)\left(1-\frac{1}{q}\right)^3
\end{aligned}$$

$$+\,mn\left(1-\frac{1}{q}\right)^2$$

From this we can see, for example, that for $m \approx n$, $m, n \to \infty$, the variance is on the order of n^2 and the standard deviation is of order n.

REFERENCES

[1] D. C. Lay. *Linear Algebra and Its Applications*, second edition. Addison-Wesley, Reading, Massachusetts, 1997.